Sporadic SARS-CoV-2 cases at the neighbourhood level in Toronto, Ontario, 2020: a spatial analysis of the early pandemic period

Lindsay Obress MSc, Olaf Berke PhD, David N. Fisman MD MPH, Ashleigh R. Tuite PhD MPH, Amy L. Greer PhD

Abstract

Background: As the largest city in Canada, Toronto has played an important role in the dynamics of SARS-CoV-2 transmission in Ontario, and the burden of disease across Toronto neighbourhoods has shown considerable heterogeneity. The purpose of this study was to investigate the spatial variation of sporadic SARS-CoV-2 cases in Toronto neighbourhoods by detecting clusters of increased risk and investigating effects of neighbourhood-level risk factors on rates.

Methods: Data on sporadic SARS-CoV-2 cases, at the neighbourhood level, for Jan. 25 to Nov. 26, 2020, were obtained from the City of Toronto COVID-19 dashboard. We used a flexibly shaped spatial scan to detect clusters of increased risk of sporadic COVID-19. We then used a generalized linear geostatistical model to investigate whether average household size, population density, dependency ratio and prevalence of low-income households were associated with sporadic SARS-CoV-2 rates.

Results: We identified 3 clusters of elevated risk of SARS-CoV-2 infection, with standardized morbidity ratios ranging from 1.59 to 2.43. The generalized linear geostatistical model found that average household size (relative risk [RR] 2.17, 95% confidence interval [CI] 1.80–2.61) and percentage of low-income households (RR 1.03, 95% CI 1.02–1.04) were significant predictors of sporadic SARS-CoV-2 cases at the neighbourhood level.

Interpretation: During the study period, 3 clusters of increased risk of sporadic SARS-CoV-2 infection were identified, and average household size and percentage of low-income households were found to be associated with sporadic SARS-CoV-2 rates at the neighbourhood level. The findings of this study can be used to target resources and create policy to address inequities that are shown through heterogeneity of SARS-CoV-2 cases at the neighbourhood level in Toronto, Ontario.

The first case of COVID-19 in Canada was reported on Jan. 25, 2020, after an individual returned to Toronto, Ontario, from Wuhan, China.[1] As the pandemic continued, Toronto remained a focal area of SARS-CoV-2 spread within Canada as the largest major city and home of Canada's busiest airport. As of Nov. 26, 2020, there were 39 914 cases of SARS-CoV-2 infection reported in Toronto with a cumulative incidence of 1220.3 cases per 100 000 population.[2,3] At that time, the cumulative incidence in the province of Ontario was 748.2 cases per 100 000 population.[4]

Sporadic cases, those without a connection to an outbreak, are important to monitor as they are indicative of the underlying community transmission. Sporadic cases differ from outbreaks that occur within specific settings such as workplaces or congregate living settings. Examining sporadic cases can be critical to understanding the dynamics of spread outside of these outbreak settings and inform decisions to improve preventive measures for the general population.

The risk and costs of a pandemic are not equal for all citizens. People with low socioeconomic status disproportionally shoulder the burden of disease in any society, and this is amplified during a global health crisis.[5] Lower socioeconomic status is associated with comorbidities linked to more severe COVID-19 and also with the conduct of essential work that cannot be done from home, such that these workers have continued to engage in in-person work throughout the pandemic.[5,6] The spatial distribution of disease can provide insight into the observed differences in disease rates across a city through examination of underlying social determinants of health and their relation to neighbourhood infection rates.

Toronto is subdivided into 140 neighbourhoods, and the burden of SARS-CoV-2 infection has been observed to vary widely across the city.[2] The goal of this study was to determine whether there are clusters of increased risk of sporadic SARS-CoV-2 infection at the neighbourhood level, to determine whether there is spatial clustering in sporadic SARS-CoV-2 rates in Toronto, and to create a generalized linear geostatistical model to investigate the effect of various risk factors on sporadic SARS-CoV-2 rates across Toronto.

Methods

Study design and setting

This study is a spatial analysis of observational data from SARS-CoV-2 cases in Toronto, by neighbourhood, reported from Jan. 25 to Nov. 26, 2020. We used neighbourhood-level SARS-CoV-2 rates to identify clusters of increased risk and investigate the effects of various area-level risk factors. We followed the Reporting of Studies Conducted Using Observational Routinely Collected Health Data (RECORD) Statement checklist when reporting the findings.[7]

Data sources

The SARS-CoV-2 case data were retrieved from the City of Toronto COVID-19 dashboard for cases reported from Jan. 25 to Nov. 26, 2020.[2] A case is defined as a confirmed or probable case of SARS-CoV-2 infection reported to Toronto Public Health through the Integrated Public Health Surveillance System and the Public Health Case and Contact Management Solution.[8] To explore the dynamics of spread at the community level, sporadic cases were selected, and outbreak-related cases were excluded. The definition of sporadic cases is "all cases that are not linked to an outbreak in general members of the population."[8]

The neighbourhood profiles and geographic boundary files were retrieved from Toronto Open Data.[9,10] Case data and neighbourhood profiles, and geographical data files were linked by neighbourhood ID numbers. Population; average household size; population density; low-income measure, after tax (LIM-AT); percentage visible minority; and population size broken down by age group were selected from the 2016 Toronto Neighbourhood Profiles as variables of interest.[9] We selected these variables as they are a subset of variables used to construct the Ontario Marginalization Index, a widely used index that encompasses various factors of marginalization and socioeconomic status but is not available at the neighbourhood level.[11] We used population by age group to create a dependency ratio calculated as the ratio of children (< 15 yr) and seniors (≥ 65 yr) to the population aged 15–64 years for each neighbourhood.[11]

Case mapping

The incidence rate of sporadic cases of SARS-CoV-2 infection reported from Jan. 25 to Nov. 26, 2020, in Toronto was mapped at the neighbourhood level. We used neighbourhood population size as the denominator to calculate the incidence rate for each neighbourhood. To account for varying population sizes across neighbourhoods, we estimated empirical Bayesian smoothed rates and visualized their spatial distribution pattern by choropleth mapping.[12] The UTM 17N projection was applied to minimize distortion of maps.[13]

Case cluster detection

We used a flexibly shaped spatial scan test to determine the locations of probable geographic clusters of elevated sporadic SARS-CoV-2 rates and estimate the standardized morbidity ratio (SMR) within identified clusters.[14] The flexibly shaped spatial scan test was selected as it allows for irregularly shaped clusters to be detected that would not be picked up by more traditional methods (e.g., circular scanning window). The spatial scan test identifies clusters by gradually scanning each neighbourhood and increasing the scanning window to a maximum cluster size. The window that attains the maximum likelihood is identified as the primary, most likely, cluster. Additional clusters may then be identified.

In this study, the maximum number of regions in a cluster was set to 14, as this represented 10% of neighbourhoods and the respective population would be still below the maximum 50% of the total population. Identifying small clusters is preferred for public health studies to allow for intervention to be applied more easily, and clusters larger than 10%–15% of the total regions are unlikely.[14]

We estimated p values to determine significance of the spatial scan test using 999 Monte Carlo simulations, where the null hypothesis is that the rate of cases within a cluster does not differ from the rate outside of the cluster.

We calculated the SMR by dividing the observed cases by the expected cases calculated in the flexibly shaped spatial scan test.[14] We excluded clusters in which the lower bound of the SMR 95% confidence interval (CI) was below 1.5, as spatial scan tests are most suitable to detect clusters with a relative risk (RR) of 1.5 and above.[15] Additionally, it was determined that an SMR above 1.5 would be of public health interest. Therefore, we excluded clusters with an SMR 95% CI that was lower than 1.5.

To determine whether case clustering (spatial dependence) was present in our data, we calculated the 2-sided Moran's I correlation coefficient using the empirical Bayesian smoothed rates, where the null hypothesis is absence of spatial correlation.[16] Queen-neighbourhood structure was used for the test, in which regions that share any border point are considered neighbours.

Generalized linear geostatistical model

To investigate SARS-CoV-2 clustering in Toronto neighbourhoods further, we built a model to examine risk factors. First, univariable Poisson regression models were used to investigate the effect of the selected risk factors on the rate of sporadic SARS-CoV-2 infection, where variables with significant p values were included in the multivariable model. To account for spatial autocorrelation, a generalized linear geostatistical model (GLGM) was fit to model the effect of average household size, population density, LIM-AT, percentage

visible minority, and dependency ratio on the number of sporadic SARS-CoV-2 cases at the neighbourhood level with population as the offset.[17] The data are centred at the centroid of each neighbourhood, and we used Euclidean distance to measure distances between neighbourhoods. The GLGM with a spherical spatial correlation structure with a Poisson family distribution was fit by penalized quasi-likelihood estimation. We assessed the model by examining the normality assumption of the standardized residuals.

Statistical analysis
We used R 4.0.2 to conduct all analyses, including generating choropleth maps, flexible scan test (smerc package), spatial clustering tests (spdep package), and fitting GLGM (MASS and GeoR packages). A significance level of 5% was used for all tests and CIs.

Ethics approval
Ethics approval for this project was not required as the data were obtained from public sources and were anonymous population-level data.

Results

The data set contained 30598 sporadic cases of SARS-CoV-2 infection in Toronto across the 140 neighbourhoods. Of these, 2.3% (704 cases) had missing postal codes and were excluded

from the analyses. Reported laboratory-confirmed case counts within a neighbourhood ranged from 27 to 1115, with empirical Bayesian smoothed rates ranging from 263.8 to 3367.8 cases per 100000 population, and with a median of 823.5 cases per 100000 population (Appendix 1, available at www.cmajopen.ca/content/10/1/E190/suppl/DC1). Rates appeared to be the highest in the northwestern regions and northeastern regions of the city and lowest in the southern and central regions (Figure 1).

Case clusters
The flexible scan test identified 3 regions of increased sporadic SARS-CoV-2 risk (Table 1, Figure 2). The primary cluster had the highest SMR of 2.43 (95% CI 2.38–2.49), meaning there is a 2.43 times higher risk within this cluster compared with the risk of sporadic SARS-CoV-2 infection within the whole city of Toronto. The SMRs of the secondary clusters were 1.59 (95% CI 1.53–1.66) and 1.70 (95% CI 1.59–1.82) (Table 1).

Moran's I test for clustering showed that spatial clustering was present, indicating there is spatial dependence in the data that must be accounted for when modelling. The value of the Moran's I coefficient was 0.676 ($p < 0.01$).

Generalized linear geostatistical model
The univariable analyses showed that all selected risk factors had an effect on sporadic SARS-CoV-2 risk at the neighbourhood level, and therefore they were all included in the
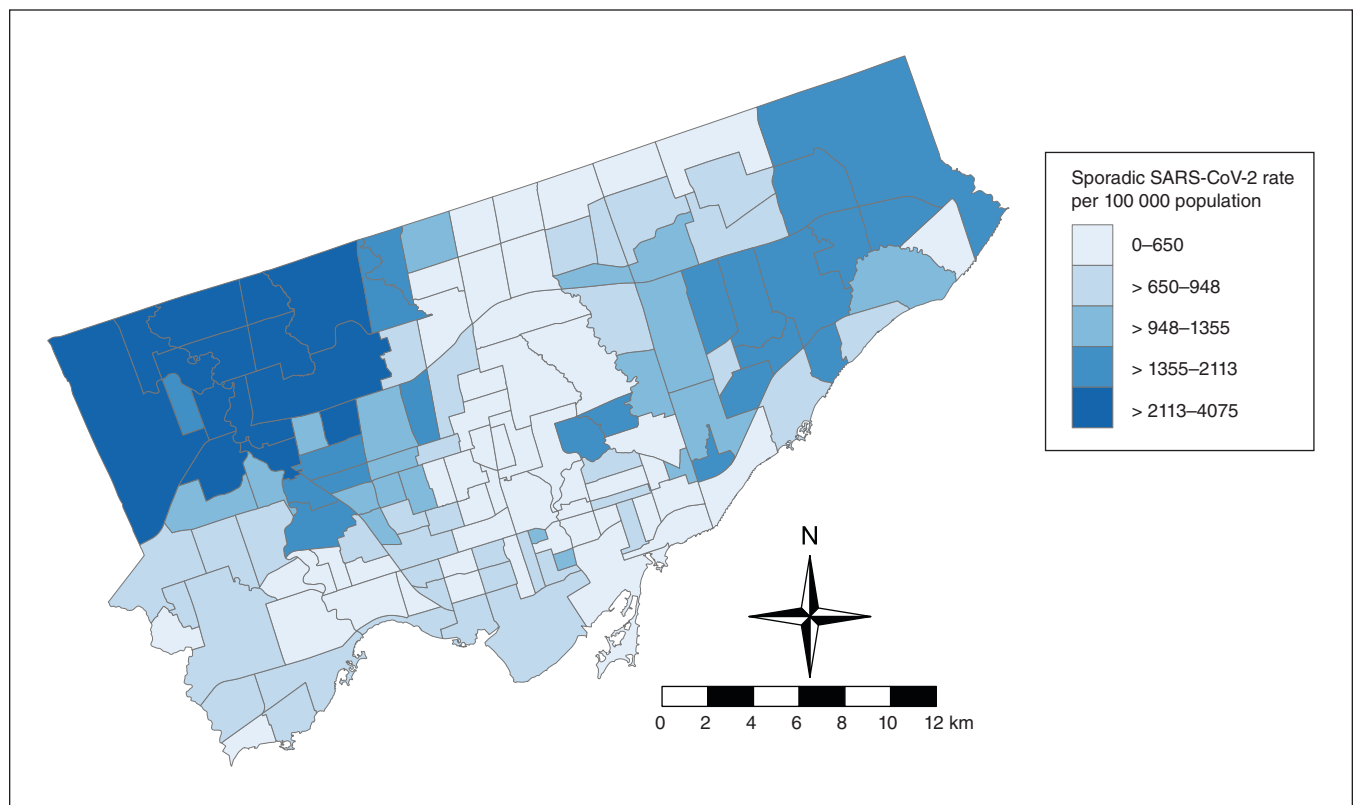

Figure 1: Map of raw sporadic SARS-CoV-2 cases per 100000 population in Toronto neighbourhoods from Jan. 25 to Nov. 26, 2020. Contains information licensed under the Open Government Licence — Toronto.

Table 1: Clusters of increased risk of sporadic SARS-CoV-2 infection in Toronto neighbourhoods

Cluster	Population	Cases	Expected	SMR (95% CI)
1	262 566	6995	2873.49	2.43 (2.38–2.49)
2	133 499	2323	1461.00	1.59 (1.53–1.66)
3	43 041	802	471.04	1.70 (1.59–1.82)

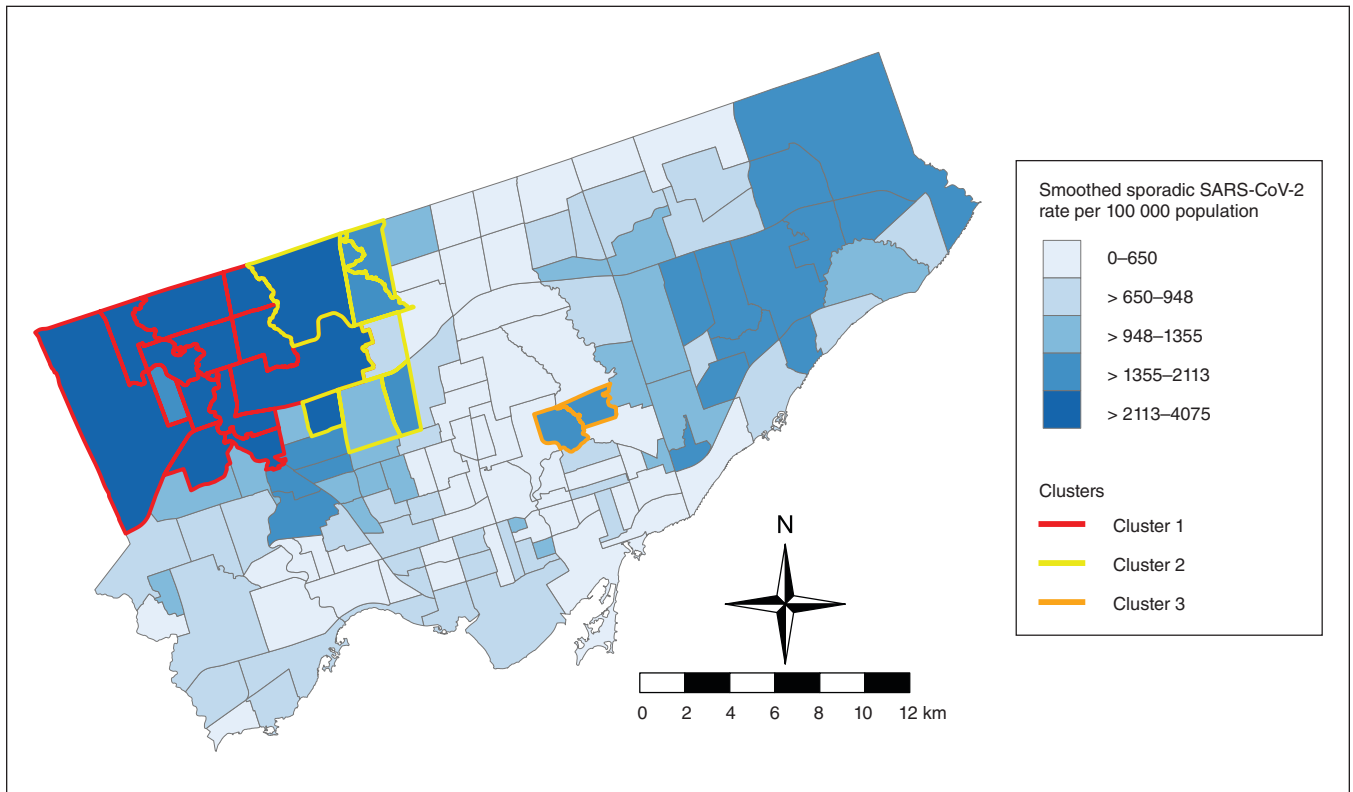Note: CI = confidence interval, SMR = standardized morbidity ratio.

Figure 2: Map showing 3 identified clusters of elevated sporadic SARS-CoV-2 risk in Toronto. Contains information licensed under the Open Government Licence — Toronto.

multivariable analysis (Table 2). A GLGM was fit, and there was a significant effect of household size and percentage of low-income households (defined by LIM-AT) on risk of sporadic SARS-CoV-2 cases. Population density, percentage visible minority and dependency ratio were not significant in the model and were removed. The percentage visible minority variable was found to be correlated with household size and percentage of low-income households in the correlation structure of the GLGM.

The final GLGM, including only average household size and percentage of low-income households, found both variables to be significant (Table 3). When average household size increases by 1, the risk of sporadic SARS-CoV-2 infection increases by 2.17 (β = 0.772, RR 2.17, p < 0.01), and a 1% increase in LIM-AT score increases risk of sporadic SARS-CoV-2 infection by 1.03 (β = 0.032, RR 1.03, p < 0.01) (Table 3). The range, the maximum distance between

centroids of neighbourhoods up to which spatial dependence is observed by the model, was 591 m. Residual analysis found no violation of the normality assumption.

Interpretation

During the study period, 3 clusters of elevated risk of sporadic SARS-CoV-2 cases were found within Toronto neighbourhoods, with SMRs ranging from 1.59 to 2.43. Although cluster 1 was identified as the primary, most likely, cluster through the spatial scan test, all clusters are of importance for public health considerations. These clusters can be identified as key areas for targeting of additional COVID-19 resources, such as pop-up testing clinics or targeted areas for vaccination.

The GLGM found that average household size and LIM-AT prevalence were associated with the rate of sporadic

Table 2: Univariable Poisson regression of the effect of each variable on sporadic SARS-CoV-2 cases in Toronto, at the neighbourhood level

Variable	Parameter estimate	Standard error	Relative risk (95% CI)
Average household size	0.795	0.014	2.21 (2.16–2.27)
Dependency ratio	1.27	0.055	3.57 (3.21–3.98)
LIM-AT	0.027	7.3×10^{-4}	1.027 (1.026–1.029)
Population density	-2.6×10^{-5}	1.4×10^{-6}	0.999974 (0.999971–0.999977)
% visible minority	0.015	2.7×10^{-4}	1.015 (1.014–1.016)

Note: CI = confidence interval; LIM-AT = low-income measure, after tax.

Table 3: Summary of generalized linear geostatistical model of sporadic SARS-CoV-2 cases in Toronto, at the neighbourhood-level

Variable	Parameter estimate	Standard error	Relative risk (95% CI)
Intercept	−7.281	0.2493	0.0007 (0.0004–0.0011)
Average household size	0.772	0.0951	2.17 (1.80–2.61)
LIM-AT	0.032	0.0048	1.03 (1.02–1.04)
Range, m	591	–	–

Note: CI = confidence interval; LIM-AT = low-income measure, after tax.

SARS-CoV-2 infection at the neighbourhood level. For average household size, when the average household size in a neighbourhood increased by 1, the risk of sporadic SARS-CoV-2 infection increased by a factor of 2.17. Additionally, as the percentage of households that fall within the low-income measure criteria increased by 1%, the risk of sporadic SARS-CoV-2 cases increased by a factor of 1.03, at the neighbourhood level. Considering the difference between the neighbourhoods with lowest LIM-AT prevalence (4.5%) and the neighbourhoods with the highest prevalence (45.5%) (Appendix 1), there was a 3.67 times higher risk of sporadic SARS-CoV-2 infection for individuals living in the area with the highest LIM-AT prevalence. The model also had a low value for range, 591 m. A range this low suggests that the spatial clustering can be explained by the risk factors included in the model and the identified clusters.

These findings align with literature linking poorer health outcomes to decreased socioeconomic status at a local level.[5] A large-scale event such as a global pandemic only widens the discrepancies between those who are more and less privileged.[5] Individuals who are of higher socioeconomic status often have jobs where they can work from home more easily than those who are of lower socioeconomic status.[6] Those of lower socioeconomic status often work in fields that have been deemed essential during a pandemic, such as health care, manufacturing and retail, and may rely on public transit to get to their place of work.[6,18]

This study provides insights into the variability observed in the spatial distribution of SARS-CoV-2 cases during a pandemic. Further studies could examine additional factors that may better characterize socioeconomic status and marginalization. For example, using the Ontario Marginalization Index could be more representative of marginalization and socioeconomic status and can be constructed using Census information; however, this was beyond the scope of this project.[11] Individual-level factors would also be of interest to examine, including occupations, ability to work from home, risk-taking behaviours, or children attending school in-person versus online. A separate research question could examine outbreak-related cases, such as in long-term care or school settings.

Limitations

Various biases and limitations may have been encountered during our analysis of these data. First, the study period includes about 9 months of the COVID-19 pandemic in 2020, and there have been a variety of changes to policy and disease dynamics (such as the emergence of new variants and development of vaccines) over the course of the pandemic. If the data were updated, results may vary owing to these factors, and future studies could use this methodology at different time points.

With regard to the data, we examined only a limited set of group-level factors and summary values. This does not often give the full picture and may miss individual variation, such as specific sex, age and race or ethnicity differences; additional variables may be of interest in future studies.

Only sporadic cases were investigated, which could be influenced by misclassification bias. For example, individuals who work in a health care setting who test positive may be deemed part of an outbreak when their infection was acquired sporadically in the community, or vice versa. There has also been found to be variation in testing rates across regions, which may also influence the number of cases being detected in neighbourhoods.

Additionally, when interpreting spatial studies, it is always important to consider the modifiable areal unit problem that occurs when studies aggregate spatial data to regions. The level of aggregation selected (in this study, the neighbourhood level) affects the interpretation of the findings; results may vary if another level of aggregation is selected (such as census tract or dissemination area).

The flexibly shaped spatial scan test has limitations, including that it is most practical for detection of small clusters; if larger clusters are to be considered, alternative methods would need to be used.[14] In addition to this, analytic choices were made that we felt best suited the data and the questions being addressed, such as the choices of a maximum cluster size of 14. Changing these cut-off points could lead to different results from those found in this study. These factors must be considered in applying these findings.

Conclusion

We found wide variation in the spatial distribution of sporadic SARS-CoV-2 incidence rates across Toronto's 140 neighbourhoods, with 3 clusters of increased risk. This variation can be at least partially explained by the risk factors that were considered in this study — that residents of areas with higher average household size and higher prevalence of low-income households had a higher risk of sporadic SARS-CoV-2 infection. Policies such as paid sick days, hotel quarantine sites and targeted vaccination strategies may help address inequities identified in this study and help prevent the spread of SARS-CoV-2.

References

1. Public Health Agency of Canada COVID-19 Surveillance and Epidemiology Team. A retrospective analysis of the start of the COVID-19 epidemic in Canada, January 15–March 12, 2020. *Can Commun Dis Rep* 2020;46:236-41.
2. COVID-19: pandemic data. Toronto: City of Toronto; updated 2021 June 1. Available: https://www.toronto.ca/home/covid-19/covid-19-latest-city-of-toronto-news/covid-19-status-of-cases-in-toronto/ (accessed 2020 Nov. 28).
3. COVID-19: epidemiologic summaries from Public Health Ontario. Toronto: Ontario Ministry of Health. Available: https://covid-19.ontario.ca/covid-19-epidemiologic-summaries-public-health-ontario (accessed 2020 Nov. 27).
4. Ontario COVID-19 Data Tool. Toronto: Public Health Ontario; 2021. Available: https://www.publichealthontario.ca/en/data-and-analysis/infectious-disease/covid-19-data-surveillance/covid-19-data-tool?tab=summary (accessed 2021 Mar. 19).
5. Singu S, Acharya A, Challagundla K, et al. Impact of social determinants of health on the emerging COVID-19 pandemic in the United States. *Front Public Health* 2020;8:406.
6. Baker MG. Nonrelocatable occupations at increased risk during pandemics: United States, 2018. *Am J Public Health* 2020;110:1126-32.
7. Benchimol EI, Smeeth L, Guttmann A, et al.; RECORD Working Committee. The reporting of studies conducted using observational routinely-collected health data (RECORD) statement. *PLoS Med* 2015;12:e1001885.
8. City of Toronto COVID-19 monitoring dashboard technical notes. Toronto: City of Toronto; updated 2021 Oct. 6:1-8. Available: https://drive.google.com/file/d/1kq0d6sSLAFt2l8BUbnofn1-SrhBPREV6/view (accessed 2020 Dec. 10).
9. Open data catalogue: neighbourhood profiles. Toronto: City of Toronto; updated 2019 Oct. 7. Available: https://open.toronto.ca/dataset/neighbourhood-profiles/ (accessed 2020 Nov. 28).
10. Open data catalogue: about neighbourhoods. Toronto: City of Toronto; updated 2021 Mar. 15. Available: https://open.toronto.ca/dataset/neighbourhoods/ (accessed 2020 Nov. 28).
11. Matheson FI, van Ingen T. 2016 Ontario marginalization index: user guide. Toronto: St. Michael's Hospital; 2018. Available: https://www.publichealthontario.ca/-/media/documents/O/2017/on-marg-userguide.pdf (accessed 2020 Nov. 28).
12. Berke O. Choropleth mapping of regional count data of *Echinococcus multilocularis* among red foxes in Lower Saxony, Germany. *Prev Vet Med* 2001;52:119-31.
13. The UTM Grid: map projections. Ottawa: Natural Resources Canada; 2021. Available: https://www.nrcan.gc.ca/earth-sciences/geography/topographic-information/maps/utm-grid-map-projections/9775 (accessed 2021 Nov. 7).
14. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geogr* 2005;4:11.
15. Aamodt G, Samuelsen SO, Skrondal A. A simulation study of three methods for detecting disease clusters. *Int J Health Geogr* 2006;5:15.
16. Assunção RM, Reis EA. A new proposal to adjust Moran's *I* for population density. *Stat Med* 1999;18:2147-62.
17. Diggle PJ, Riberio PJ Jr. Generalized linear models for geostatistical data. In: *Model-based geostatistics*. New York: Springer; 2007:79-98.
18. Sy KTL, Martinez ME, Rader B, et al. Socioeconomic disparities in subway use and COVID-19 outcomes in New York City. *Am J Epidemiol* 2021;190:1234-42.

Affiliations: Department of Population Medicine (Obress, Berke, Greer), University of Guelph, Guelph, Ont.; Dalla Lana School of Public Health (Fisman, Tuite, Greer), University of Toronto, Toronto, Ont.